Joint Modeling of Call and Put Implied Volatility

Katja Ahoniemi
Helsinki School of Economics, FDPE, and HECER

and

Markku Lanne
University of Helsinki, RUESG, and HECER

Joint Modeling of Call and Put Implied Volatility*

Abstract

This paper exploits the fact that implied volatilities calculated from identical call and put options have often been empirically found to differ, although they should be equal in theory. We propose a new bivariate mixture multiplicative error model and show that it is a good fit to Nikkei 225 index call and put option implied volatility (IV). A good model fit requires two mixture components in the model, allowing for different mean equations and error distributions for calmer and more volatile days. Forecast evaluation indicates that in addition to jointly modeling the time series of call and put IV, cross effects should be added to the model: put-side implied volatility helps forecast call-side IV, and vice versa. Impulse response functions show that the IV derived from put options recovers faster from shocks, and the effect of shocks lasts for up to six weeks.

JEL Classification: C32, C53, G13

Keywords: Implied Volatility, Option Markets, Multiplicative Error Models, Forecasting.

Katja Ahoniemi   Markku Lanne

Department of Economics   Department of Economics
Helsinki School of Economics   University of Helsinki
P.O. Box 1210   P.O. Box 17 (Arkadiankatu 7)
FI-00101 Helsinki   FI-00014 University of Helsinki
FINLAND   FINLAND

e-mail: katja.ahoniemi@hse.fi   e-mail: markku.lanne@helsinki.fi

1   Introduction

In theory, the implied volatilities derived from a call option and a put option with the same underlying asset, strike price, and expiration date should be equal - both reflect the market's expectation of the volatility of the returns of the underlying asset during the remaining life of the two options. However, it has been empirically observed that when call and put implied volatilities (IV) are backed out of option prices using an option pricing formula, they often deviate from each other.

The reason behind the inequality of put and call implied volatilities may lie in the different demand structure for calls and puts. There is an inherent demand for put options that does not exist for similar calls, as institutional investors buy puts regularly for purposes of portfolio insurance. There are often no market participants looking to sell the same options to offset this demand, meaning that prices may need to be bid up high enough for market makers to be willing to become counterparties to the deals. With no market imperfections such as transaction costs or other frictions present, option prices should always be determined by no-arbitrage conditions, making implied volatilities of identical call and put options the same. However, in real-world markets the presence of imperfections may allow option prices to depart from no-arbitrage bounds if there is, for example, an imbalance between supply and demand in the market. References to existing literature and more details on this topic are provided in Section 2.

Despite the fact that call and put-side implied volatilities differ, they must be tightly linked to one another at all times - after all, they both represent the same market expectation, and the driving forces behind their values are common. Therefore, it can be argued that there is potential value added in jointly modeling time series of implied volatilities, one derived from call option prices and the other from put option prices. Further, the interactions between the two variables can be studied with cross effects, i.e. allowing call IV to depend on lagged values of put IV, and vice versa.

The modeling of IV provides a valuable addition to the extensive literature on volatility modeling. IV is truly a forward-looking measure: implied volatility is the market's expectation of the volatility in the returns of an option's underlying asset during the remaining life of the option in question. In contrast, other volatility estimates are based on historical prices. Examples of IV modeling literature include Ahoniemi (2006), who finds that there is some predictability in the direction of change of the VIX Volatility Index, an index of the IV of S&P 500 index options. Dennis et al. (2006) find that daily innovations in the VIX Volatility Index contain very reliable incremental information about the future volatility of the S&P 100 index.[1] Other studies that attempt to forecast IV or utilize the information contained in IV to trade in option markets include Harvey and Whaley (1992), Noh et al. (1994), and Poon and Pope (2000). Reliable forecasts of implied volatility can benefit option traders, but many other market participants as well: all investors with risk management concerns can benefit from accurate forecasts of future volatility.

The implied volatility data used in this study are calculated separately from call and put options on the Japanese Nikkei 225 index. Separate time series for call and put-side IV offer a natural application for the bivariate multiplicative model presented below. In their analysis of implied volatilities of options on the S&P 500 index, the FTSE 100

---

[1]The data set in Dennis et al. (2006) ends at the end of 1995, when options on the S&P 100 index were used to calculate the value of the VIX. The Chicago Board Options Exchange has since switched to S&P 500 options.

index, and the Nikkei 225 index, Mo and Wu (2007) find that U.S. and UK implied volatilities are more correlated with each other than with Japanese implied volatilities, indicating that the Japanese market exhibits more country-specific movements. Therefore, it is interesting to analyze the Japanese option market and its implied volatility in this context, as investors may be presented with possibilities in the Japanese index option market that are not available elsewhere. Mo and Wu (2007) also report that the implied volatility skew is flatter in Japan than in the U.S. or UK markets. They conclude that in Japan, the risk premium for global return risks is smaller than in the other two countries. The developments in the Japanese stock market during the late 1990s in particular are very different from Western markets, with prices declining persistently in Japan. This characteristic also makes the Japanese market unique. Mo and Wu (2007) observe that out-of-the-money calls have relatively higher IVs in Japan, as investors there expect a recovery after many years of economic downturn. Investors in Japan seem to price more heavily against volatility increases than against market crashes.

In this paper, we introduce a new bivariate multiplicative error model (MEM). MEM models have gained ground in recent years due to the increasing interest in modeling non-negative time series in financial market research.[2] The use of MEM models does not require logarithms to be taken of the data, allowing for the direct modeling of variables such as the duration between trades, the bid-ask spread, volume, and volatility. Recent papers that successfully employ multiplicative error modeling in volatility applications include Engle and Gallo (2006), Lanne (2006, 2007), and Ahoniemi (2007). Lanne (2006) finds that the gamma distribution is well suited for the multiplicative modeling of the realized volatility of two exchange rate series, and Ahoniemi (2007), using the same data set as in the present study, finds that MEM models together with a gamma error distribution are a good fit to data on Nikkei 225 index implied volatility. All the above-mentioned MEM applications consider univariate models, but Cipollini et al. (2006) build a multivariate multiplicative error model using copula functions instead of directly employing a multivariate distribution. In our application, we use a bivariate gamma distribution to model the residuals.

Our results show that it is indeed useful to jointly model call and put implied volatilities. The chosen mixture bivariate model with a gamma error distribution is a good fit to the data, as shown by coefficient significance and diagnostic checks. The addition of lagged cross effects turns out to be important for one-step-ahead daily forecast performance. Our model correctly forecasts the direction of change in IV on over 70% of trading days in an out-of-sample analysis. Impulse response functions are also calculated, and they reveal that there is considerable persistence in the data: shocks do not fully disappear until thirty trading days elapse. Also, put-side IV recovers more quickly from shocks than call-side IV, indicating that the market for put options may price more efficiently due to larger demand and trading volumes.

This paper proceeds as follows. Section 2 discusses the differences in the markets for call and put options in more detail. Section 3 describes the bivariate mixture multiplicative error model estimated in this paper. Section 4 presents the data, model estimation results, and diagnostic checks of the chosen model specification. Impulse response functions are discussed in Section 5, and forecasts are evaluated in Section 6. Section 7 concludes.

---

[2]A special case of multiplicative error models is the autoregressive conditional duration (ACD) model, for which an abundant literature has emerged over the past ten years.

2 The Markets for Call and Put Options

There is an abundance of literature investigating the differences in the markets for call and put options. Bollen and Whaley (2004) have documented that put options account for 55 % of trades in S&P 500 index options, and that the level of implied volatility calculated from at-the-money (ATM) options on the S&P 500 index is largely driven by the demand for ATM index puts. Buraschi and Jackwerth (2001), using an earlier data set of S&P 500 index options, report that put volumes are around three times higher than call volumes. There is also evidence that out-of-the-money (OTM) puts in particular can be overpriced, at least part of the time (Bates (1991), Dumas et al. (1998), Bollen and Whaley (2004)). Garleanu et al. (2006) document that end users (non-market makers) of options have a net long position in S&P 500 index puts, and that net demand for low-strike options (such as OTM puts) is higher than the demand for high-strike options. The results of Chan et al. (2004) from Hang Seng Index options in Hong Kong are similar to those of Bollen and Whaley (2004) in that net buying pressure is more correlated with the change in implied volatility of OTM put options than in-the-money put options. Also, trading in Hang Seng Index puts determines the shape of the volatility smile to a greater degree than trading in calls. If OTM puts are consistently overpriced, investors who write such options could earn excess returns (empirical evidence in support of this is provided in e.g. Bollen and Whaley (2004)). On the other hand, Jackwerth (2000) finds that it is more profitable to sell ATM puts than OTM puts in the S&P 500 index option market. Fleming (1999) compares ATM S&P 100 index calls and puts, and finds that selling puts is more profitable than selling calls.

Further evidence on different market mechanisms for calls and puts is provided by Rubinstein (1994), who notes that after the stock market crash of October 1987, prices of OTM puts were driven upwards, changing the volatility smile into the now-observed volatility skew. He hypothesizes that the crash led to OTM puts being more highly valued in the eyes of investors. Fleming (1999) observes that institutional buying pressure rose dramatically after the '87 crash. Ederington and Guan (2002) also remark that the volatility smile may be caused in part by hedging pressures which drive up the prices of puts with low strike prices. They point out that this notion is supported by both trading volume evidence and the fact that in equity markets, implied volatilities calculated from options with low strike prices have been found to be higher than actual volatilities. Das and Uppal (2004) note that downside jumps in international equity markets tend to occur at the same time. Mo and Wu (2007) also report that large downside moves are more likely to be global rather than country-specific movements. As a consequence of this, investors cannot avoid drops in portfolio value by diversifying internationally. This then creates additional pressure to acquire portfolio insurance from put options, driving up their prices.

Even if the demand for a put option causes its price (and implied volatility) to rise, no-arbitrage conditions should ensure that the price of a call option with the same strike price and maturity date yields an implied volatility that is equal to the one derived from the put counterpart. But as Fleming (1999) writes,

> ..., transaction costs and other market imperfections can allow option prices to deviate from their "true" values without signaling arbitrage opportunities.

The possibility that option prices can depart from no-arbitrage bounds, thus allowing

call and put IV to differ, has been documented numerous times in earlier work. Hentschel (2003) points out that noise and errors in option prices stemming from fixed tick sizes, bid-ask spreads, and non-synchronous trading can contribute to miscalculated implied volatilities, and to the volatility smile. Garleanu et al. (2006) develop a model for option prices that allows for departures from no-arbitrage bounds. These arise from the inability of market makers to perfectly hedge their positions at all times, which in turn allows option demand to affect option prices. Empirical evidence lends support to this theory: market makers require a premium for delivering index options. Even market makers cannot fully hedge their exposures due to issues such as transaction costs, the indivisibility of securities, and the impossibility of executing rebalancing trades continuously (Figlewski (1989)), and capital requirements and sensitivity to risk (Shleifer and Vishny (1997)). When market makers face unhedgeable risk, they must be compensated through option prices for bearing this risk. In fact, Garleanu et al. (2006) find that after periods of dealer losses, the prices of options are even more sensitive to demand. Other impediments to arbitrage include the fact that a stock index portfolio is difficult and costly to trade, but if an investor uses futures, she must bear basis and possibly tracking risk (Fleming (1999)): spot and futures prices may not move hand-in-hand at all times, and the underlying asset of the futures contracts may not be identical to the asset being hedged. Liu and Longstaff (2004) demonstrate that it can often be optimal to underinvest in arbitrage opportunities, as mark-to-market losses can be considerable before the values of the assets involved in the trade converge to the values that eventually produce profits to the arbitrageur. When it is suboptimal to fully take advantage of an arbitrage opportunity, there is no reason why the arbitrage could not persist for even a lengthy amount of time. Bollen and Whaley (2004), in their analysis of the S&P 500 option market, find support for the hypothesis that limits to arbitrage allow the demand for options to affect implied volatility.

3 The Model

In this section, we present the bivariate mixture multiplicative error model (BVMEM) that will be used to model the two time series of implied volatilities described in Section 4. Consider the following bivariate model

$$\mathbf{v}_t = \boldsymbol{\mu}_t \boldsymbol{\varepsilon}_t, \qquad t = 1, 2, ..., T,$$

where the conditional mean

$$\boldsymbol{\mu}_t = \left(\begin{array}{c} \mu_{1t} \\ \mu_{2t} \end{array} \right) = \left(\begin{array}{c} \omega_1 + \sum_{i=1}^{q_1} \alpha_{1i} v_{1,t-i} + \sum_{j=1}^{p_1} \beta_{1j} \mu_{1,t-j} \\ \omega_2 + \sum_{i=1}^{q_2} \alpha_{2i} v_{2,t-i} + \sum_{j=1}^{p_2} \beta_{2j} \mu_{2,t-j} \end{array} \right)$$

and $\boldsymbol{\varepsilon}_t$ is a stochastic positive-valued error term such that $E\left(\boldsymbol{\varepsilon}_t | \mathfrak{F}_{t-1}\right) = \mathbf{1}$ with $\mathfrak{F}_{t-1} = \{\mathbf{v}_{t-j}, j \geq 1\}$. In what follows, this specification will be called the BVMEM$(p_1, q_1; p_2, q_2)$ model. As the conditional mean equations of the model are essentially the same as the conditional variance equations in the GARCH model in structure, the constraints on parameter values that guarantee positivity in GARCH models also apply to each of the equations of the BVMEM model. As outlined in Nelson and Cao (1992), the parameter values in a first-order model must all be non-negative. In a higher-order model, positivity of all parameters is not necessarily required. For example, in a model with $p_i = 1$ and $q_i = 2$, $i = 1, 2$, the constraints are $\omega_i \geq 0$, $\alpha_{i1} \geq 0$, $0 \leq \beta_i < 1$, and $\beta_1 \alpha_{i1} + \alpha_{i2} \geq 0$. It

should be noted that this basic conditional mean specification must often be augmented with elements such as cross effects between the variables and seasonality effects. In these cases, one must ensure that positivity continues to be guaranteed. For example, if the coefficients for lagged cross terms are positive, no problems in achieving positivity arise.

The multiplicative structure of the model was suggested for volatility modeling in the univariate case by Engle (2002), who proposed using the exponential distribution. However, the gamma distribution nests, among others, the exponential distribution, and is therefore more general. Also, the findings of Lanne (2006, 2007) and Ahoniemi (2007) lend support to the gamma distribution.

The error term ε_t is assumed to follow a bivariate gamma distribution, which is a natural extension of the univariate gamma distribution used in previous literature (Lanne (2006, 2007) and Ahoniemi (2007)). Of the numerous bivariate distributions having gamma marginals, the specification suggested by Nagao and Kadoya (1970) is considered (for a discussion on alternative bivariate gamma densities, see Yue et al. (2001)). This particular specification is quite tractable and thus well suited for our purposes. Collecting the parameters into vector $\boldsymbol{\theta} = (\tau_1, \tau_2, \lambda, \rho)$, the density function can be written as

$$f_{\varepsilon_1,\varepsilon_2}\left(\varepsilon_{1t}, \varepsilon_{2t}; \boldsymbol{\theta}\right) = \frac{(\tau_1 \tau_2)^{(\lambda+1)/2} (\varepsilon_{1t}\varepsilon_{2t})^{(\lambda-1)/2} \exp\left\{-\frac{\tau_1 \varepsilon_{1t} + \tau_2 \varepsilon_{2t}}{1-\rho}\right\}}{\Gamma(\lambda)(1-\rho)\rho^{(\lambda-1)/2}} I_{\lambda-1}\left(\frac{2\sqrt{\tau_1 \tau_2 \rho \varepsilon_{1t}\varepsilon_{2t}}}{1-\rho}\right),$$

where $\Gamma(\cdot)$ is the gamma function, ρ is the Pearson product-moment correlation coefficient, and $I_{\lambda-1}(\cdot)$ is the modified Bessel function of the first kind. The marginal error distributions have distinct scale parameters τ_1 and τ_2, but the shape parameter, λ, is the same for both. However, since the error term needs to have mean unity, we impose the restrictions that the shape parameters are the reciprocals of the scale parameters, i.e., $\lambda = 1/\tau_1$ and $\lambda = 1/\tau_2$, indicating that $\tau_1 = \tau_2 = 1/\lambda$. In other words, we will also restrict the scale parameters to be equal. This is not likely to be very restrictive in our application, as earlier evidence based on univariate models in Ahoniemi (2007) indicates that the shape and scale parameters for the time series used in this study, the implied volatilities of Nikkei 225 call and put options, are very similar.

Incorporating the restrictions discussed above and using the change of variable theorem, the conditional density function of $\mathbf{v}_t = (v_{1t}, v_{2t})'$ is obtained as

$$
\begin{aligned}
f_{t-1}\left(v_{1t}, v_{2t}; \boldsymbol{\theta}\right) &= f_{\varepsilon_1,\varepsilon_2}\left(v_{1t}\mu_{1t}^{-1}, v_{2t}\mu_{2t}^{-1}\right)\mu_{1t}^{-1}\mu_{2t}^{-1} \qquad (1)\\
&= \frac{\lambda^{(\lambda+1)}\left[v_{1t}v_{2t}\mu_{1t}^{-1}\mu_{2t}^{-1}\right]^{(\lambda-1)/2}\exp\left\{-\frac{\lambda\left(v_{1t}\mu_{1t}^{-1}+v_{2t}\mu_{2t}^{-1}\right)}{1-\rho}\right\}}{\Gamma(\lambda)(1-\rho)\rho^{(\lambda-1)/2}} \times \\
&\quad I_{\lambda-1}\left(\frac{2\lambda\sqrt{\rho v_{1t}v_{2t}\mu_{1t}^{-1}\mu_{2t}^{-1}}}{1-\rho}\right)\mu_{1t}^{-1}\mu_{2t}^{-1}.
\end{aligned}
$$

Consequently, the conditional log-likelihood function can be written as[3]

$$l_T(\boldsymbol{\theta}) = \sum_{t=1}^{T} l_{t-1}(\boldsymbol{\theta}) = \sum_{t=1}^{T} \ln\left[f_{t-1}(v_{1t}, v_{2t}; \boldsymbol{\theta})\right],$$

and the model can be estimated with the method of maximum likelihood (ML) in a straightforward manner. Although the gamma distribution is quite flexible in describing the dynamics of implied volatilities, in our empirical application it turned out to be inadequate as such. In particular, it failed to capture the strong persistence in the implied volatility time series. As an extension, we consider a mixture specification that allows for the fact that financial markets experience different types of regimes, alternating between calm and more volatile periods of time. Different parameter values can be assumed to better describe periods of larger shocks compared with periods of smaller shocks, and error terms are allowed to come from two gamma distributions whose shape and scale parameters can differ. Earlier evidence from Lanne (2006) and Ahoniemi (2007) indicates that the use of a mixture specification improves the fit of a multiplicative model as well as the forecasts obtained from the models.

We will assume that the error term $\boldsymbol{\varepsilon}_t$ is a mixture of of $\boldsymbol{\varepsilon}_t^{(1)}$ and $\boldsymbol{\varepsilon}_t^{(2)}$ with mixing probability π, and that $\boldsymbol{\varepsilon}_t^{(1)}$ and $\boldsymbol{\varepsilon}_t^{(2)}$ follow the bivariate gamma distribution with parameter vectors $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$, respectively. In other words, the error term is $\boldsymbol{\varepsilon}_t^{(1)}$ with probability π and $\boldsymbol{\varepsilon}_t^{(2)}$ with probability $1 - \pi$ ($0 < \pi < 1$). The model based on this assumption will subsequently be called the mixture-BVMEM model. The conditional log-likelihood function becomes

$$l_T(\boldsymbol{\theta}) = \sum_{t=1}^{T} l_{t-1}(\boldsymbol{\theta}) = \sum_{t=1}^{T} \ln\left[\pi f_{t-1}^{(1)}(v_{1t}, v_{2t}; \boldsymbol{\theta}_1) + (1 - \pi) f_{t-1}^{(2)}(v_{1t}, v_{2t}; \boldsymbol{\theta}_2)\right],$$

where $f_{t-1}^{(1)}(v_{1t}, v_{2t}; \boldsymbol{\theta}_1)$ and $f_{t-1}^{(2)}(v_{1t}, v_{2t}; \boldsymbol{\theta}_2)$ are given by (1) with $\boldsymbol{\theta}$ replaced by $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$, respectively.

Assuming that v_t is stationary and ergodic, it is reasonable to apply standard asymptotic results in statistical inference. In particular, approximate standard errors can be obtained from the diagonal elements of the matrix $-\left[\partial^2 l_T(\hat{\theta})/\partial\theta\partial\theta'\right]^{-1}$, where $\hat{\theta}$ denotes the ML estimate of $\boldsymbol{\theta}$. Similarly, Wald and likelihood ratio (LR) tests for general hypotheses will have the conventional asymptotic χ^2 null distributions. Note, however, that hypotheses restricting the number of mixture components do not have the usual χ^2 distributions due to the problem of unidentified parameters (see e.g. Davies (1977)). We will not attempt such tests, but assume throughout that there are two mixture components. The adequacy of the assumption will be verified by means of diagnostic procedures (see Section 4.3).

---

[3]Specifically, for observation t,

$$
\begin{aligned}
l_{t-1}(\boldsymbol{\theta}) &= (\lambda + 1)\ln(\lambda) + \frac{1}{2}(\lambda - 1)\left[\ln(v_{1t}) + \ln(v_{2t}) - \ln(\mu_{1t}) - \ln(\mu_{2t})\right] \\
&\quad - \frac{\lambda\left(v_{1t}\mu_{1t}^{-1} + v_{2t}\mu_{2t}^{-1}\right)}{1 - \rho} - \ln[\Gamma(\lambda)] - \ln(1 - \rho) - \frac{1}{2}(\lambda - 1)\ln(\rho) \\
&\quad + \ln\left[I_{\lambda-1}\left(\frac{2\lambda\sqrt{\rho v_{1t} v_{2t}\mu_{1t}^{-1}\mu_{2t}^{-1}}}{1 - \rho}\right)\right] - \ln(\mu_{1t}) - \ln(\mu_{2t}).
\end{aligned}
$$

4 Estimation Results

4.1 Data

The data set in this study covers 3,194 daily closing observations from the period 1.1.1992 - 31.12.2004, and was obtained from Bloomberg Professional Service (see Figure 1). The first eleven years of the full sample, or 1.1.1992 - 31.12.2002, comprise the in-sample of 2,708 observations. The final two years, 2003 and 2004, are left as the 486-day out-of-sample to be used for forecast evaluation.

The call-side (put-side) implied volatility time series is calculated as an unweighted average of Black-Scholes implied volatilities from two nearest-to-the-money call (put) options from the nearest maturity date. Rollover to the next maturity occurs two calendar weeks prior to expiration in order to avoid possibly erratic behavior in IV close to option expiration. ATM options are typically used to estimate the market's expected volatility for the remainder of the option's maturity, as trading volumes are usually high for ATM options. Also, ATM options have the highest sensitivity to volatility.
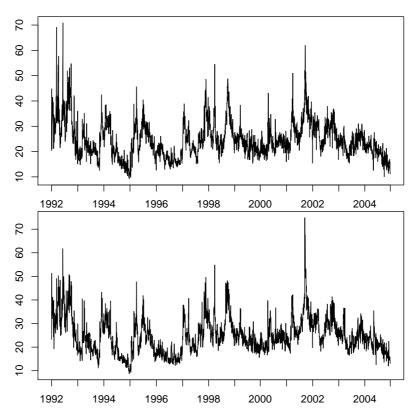
Figure 1: Nikkei 225 index call implied volatility (upper panel) and put implied volatility (lower panel) 1.1.1992 - 31.12.2004.

Table 1 provides descriptive statistics on both the call-side IVs (NIKC) and put-side IVs (NIKP). The average level of put-side implied volatility is higher in the sample of this study, a phenomenon which has also been documented in the U.S. markets by Harvey and Whaley (1992).

	NIKC	*NIKP*
Maximum	70.84	74.87
Minimum	9.26	8.80
Mean	24.68	24.82
Median	23.42	23.84
Standard deviation	7.07	7.41
Skewness	1.10	0.94
Excess kurtosis	2.42	1.79

Table 1: Descriptive statistics for NIKC and NIKP for the full sample of 1.1.1992 - 31.12.2004.

4.2 Model Estimation

Given the clear linkages between the implied volatilities of call and put options on the same underlying asset outlined above, call-side (put-side) IV can be expected to be a significant predictor of future put-side (call-side) IV. Therefore, the model presented in Section 3 is augmented with lagged cross terms, so that call (put) implied volatility depends on its own history as well as on the history of put (call) implied volatility. Bollen and Whaley (2004) find that in the U.S. market, the demand for ATM index puts drives both the changes in ATM put implied volatility and the changes in ATM call implied volatility. Therefore, we expect that for our Japanese implied volatility data, lagged put IV will be more significant in explaining call IV than lagged call IV will be in explaining put IV.

Dummy variables for Friday effects of put-side IV are also added due to the improvement in diagnostics achieved after the addition (see Section 4.3 for more details on diagnostic checks). The level of IV is lowest on Fridays for both call and put options,[4] but trading volumes are highest on Fridays. An analysis of trading volumes of close-to-the-money, near-term maturity call and put options on the Nikkei 225 index reveals that during the two-year out-of-sample period used in this study, put options account for 52.0 % of trading volume (measured with number of contracts traded). The share of puts is lowest on Mondays (50.4%) and largest on Fridays (53.6%).

Findings similar to ours concerning weekly seasonality have been reported in previous studies. Peña et al. (1999) find that in the Spanish stock index market, the curvature of the volatility smile at the beginning of the week is statistically significantly different from the smile at the end of the week. Lehmann and Modest (1994) report that trading volumes on the Tokyo Stock Exchange are substantially lower on Mondays than on other days of the week. They hypothesize that this is due to reduced demand by liquidity traders due to the risk of increased information asymmetry after the weekend. Also, bid-ask spreads are largest on Mondays, making transaction costs highest at the start of the week. The significance of trading volumes is highlighted by Mayhew and Stivers (2003), who find that implied volatility performs well when forecasting individual stock return volatility, but only for those stocks whose options have relatively high trading volumes.

In order to take cross effects and the observed seasonal variation into account, we need to modify the basic model presented in Section 3. Let μ_{mt} denote the conditional mean of mixture component m ($m = 1, 2$), and $\mu_{mt} = (\mu_{mt}^C, \mu_{mt}^P)'$, where μ_{mt}^C and μ_{mt}^P

---

[4]The level of IV is highest on Mondays. However, dummies for Monday effects were not statistically significant.

are the conditional means of the call and put implied volatilities, respectively.

The specifications of the conditional means are

$$\mu_{mt}^C = \omega_m^C + \sum_{i=1}^{q_C} \alpha_{mi}^C v_{C,t-i} + \sum_{i=1}^{r_C} \psi_{mi}^C v_{P,t-i} + \sum_{i=1}^{s_C} \delta_{mi}^{CP} D_i v_{P,t-i} + \sum_{j=1}^{p_C} \beta_{mj}^C \mu_{m,t-j}^C$$

and

$$\mu_{mt}^P = \omega_m^P + \sum_{i=1}^{q_P} \alpha_{mi}^P v_{P,t-i} + \sum_{i=1}^{r_P} \psi_{mi}^P v_{C,t-i} + \sum_{i=1}^{s_P} \delta_{mi}^{PP} D_i v_{P,t-i} + \sum_{j=1}^{p_P} \beta_{mj}^P \mu_{m,t-j}^P$$

where the ψ's are the coefficients for lagged cross terms, and D_i receives the value of 1 on Fridays, and zero otherwise. As mentioned above, the dummy variable in both the call and put mean equations is for put-side Friday effects (coefficients δ_{mi}^{CP} and δ_{mi}^{PP}). This specification is later referred to as the unrestricted model.

In order to fully understand the value of including cross effects between NIKC and NIKP in the model, an alternative specification with no cross terms was also estimated. In this model, dummies for Friday effects are also included, but due to the elimination of cross effects, the dummy in the equation for NIKC captures the Friday effect of call-side, not put-side, implied volatility. In the second model specification, or the restricted model,

$$\mu_{mt}^C = \omega_m^C + \sum_{i=1}^{q_C} \alpha_{mi}^C v_{C,t-i} + \sum_{i=1}^{s_C} \delta_{mi}^{CC} D_i v_{C,t-i} + \sum_{j=1}^{p_C} \beta_{mj}^C \mu_{m,t-j}^C$$

and

$$\mu_{mt}^P = \omega_m^P + \sum_{i=1}^{q_P} \alpha_{mi}^P v_{P,t-i} + \sum_{i=1}^{s_P} \delta_{mi}^{PP} D_i v_{P,t-i} + \sum_{j=1}^{p_P} \beta_{mj}^P \mu_{m,t-j}^P.$$

The estimation results for both the unrestricted and the restricted model are presented in Table 2. The parameter values for all ω's, α's and β's meet the Nelson and Cao (1992) constraints discussed in Section 3. Also, the coefficients of cross terms (ψ's) and dummies (δ's) are positive, so positivity is guaranteed in the model. Compared with the full version of the unrestricted model, β_{11}^C, ω_1^P, and α_{12}^P are constrained to be equal to zero, which is validated by a likelihood ratio test with p-value 0.273.[5]

The probability parameter π is quite high for the unrestricted model, close to 0.92. Therefore, the second regime, which displays larger shocks, occurs on only some eight percent of the trading days in the in-sample. The estimated shape (and scale) parameters of the error distribution differ considerably between the two regimes, with residuals more dispersed in the second regime. Figure 2 shows the joint error density of the unrestricted model with the parameters estimated for the first regime, while the error density for the second regime is depicted in Figure 3. It should be noted that the scale of the z-axis is different in the two figures. The errors are much more tightly concentrated around unity in the first, more commonly observed, regime, whereas the tail area is emphasized in the second regime.

---

[5]The model originally included six dummies: both first-regime equations had Friday-effect dummies for the intercept, own lagged value, and the lagged value of the other variable. Only the put-side Friday effects were statistically significant, and p-values from likelihood ratio tests validated the constraining of the other dummies to zero.

	Unrestricted Model	Restricted Model
Log likelihood	-12370.0	-12653.4
π	0.919** (0.012)	0.883** (0.018)
λ_1	126.594** (3.745)	127.3036** (4.614)
ρ_1	0.094** (0.027)	0.019 (0.033)
ω_1^C	1.264** (0.164)	0.298** (0.085)
α_{11}^C	0.514** (0.020)	0.617** (0.025)
α_{12}^C	0.104** (0.019)	-0.223** (0.048)
ψ_{11}^C	0.321** (0.017)	-
δ_{11}^{CC}	-	0.043** (0.006)
δ_{11}^{CP}	0.045** (0.005)	-
β_{11}^C	-	0.584** (0.046)
ω_1^P	-	0.254** (0.077)
α_{11}^P	0.529** (0.020)	0.611** (0.023)
α_{12}^P	-	-0.215** (0.053)
ψ_{11}^P	0.247** (0.018)	-
δ_{11}^{PP}	0.043** (0.005)	0.046** (0.006)
β_{11}^P	0.210** (0.026)	0.581** (0.048)
λ_2	20.043** (2.003)	23.413** (2.288)
ρ_2	0.360** (0.063)	0.378** (0.058)
ω_2^C	1.401 (0.772)	0.718 (0.368)
α_{21}^C	0.185** (0.070)	0.216** (0.044)
ψ_{21}^C	0.150 (0.101)	-
β_{21}^C	0.627** (0.108)	0.769** (0.048)
ω_2^P	0.835 (0.688)	0.933* (0.403)
α_{21}^P	0.244** (0.087)	0.270** (0.050)
ψ_{21}^P	0.146 (0.090)	-
β_{21}^P	0.612** (0.111)	0.717** (0.053)

Table 2: Estimation results for the BVMEM model. Standard errors calculated from the final Hessian matrix are given in parentheses. (**) indicates statistical significance at the one-percent level, and (*) at the five-percent level.
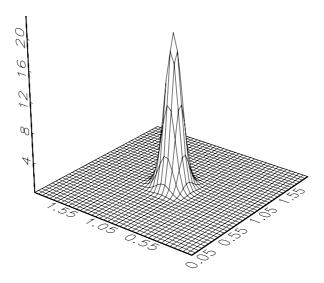
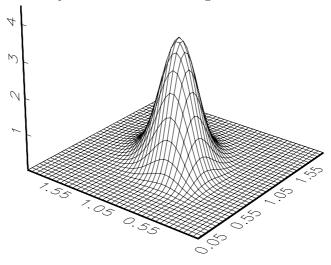Figure 2: Density of residuals in the first regime of the unrestricted model.

Figure 3: Density of residuals in the second regime of the unrestricted model.

The correlation of errors, or ρ, is higher in the second regime, making changes in call and put IV more correlated when volatility is high. This is also clearly visible in Figures 2 and 3. The coefficients of cross terms ψ are significant at the one-percent level in the first, more common regime, and jointly significant in the second regime (p-value from LR test equal to 0.007). The coefficients of the cross terms are higher in the first regime, making the cross effects more pronounced. In other words, the cross effects are smaller when volatility is high. For both regimes, the effect put-side IV has on call-side IV is larger than the effect call IV has on put IV, although the difference in coefficients is quite small in the second regime. Friday dummies for the first lag of put IV are also significant and positive, indicating that the effect of the lagged put IV is larger on Fridays, when trading volumes are highest. Values of intercepts are higher in the second regime, consistent with the notion that this regime occurs on days when shocks are larger. The clearly greater β's in the second regime indicate higher persistence in that regime. This can be interpreted as a sign that once the second regime is entered, it is likely that large shocks persist, i.e. there is volatility clustering present.

As we are interested in seeing the relevance of cross terms for forecast performance, we also present the results for the restricted model without these cross effects. It should be noted that the null hypothesis of all coefficients of cross terms equal to zero is rejected by an LR test at all reasonable significance levels. In the restricted model, the estimate of π is smaller than in the unrestricted model, but the first regime remains clearly more prevalent. The parameters of the error distribution are very similar, but the correlation of the residuals is lower in the first regime than it was with the unrestricted model. One notable difference to the parameter values of the unrestricted model is that the coefficients of the second lags are both significant in the first regime, and have a negative sign. This suggests that the exclusion of cross effects results in biased estimates of these parameters. The dummies for Friday effects are significant, indicating that the data behaves somewhat differently when the trading volume is at its highest. As the shape (and scale) parameters of the error distribution that are estimated for the restricted model are very close in value to those for the unrestricted model, the graphs for error densities are qualitatively similar as those in Figures 2 and 3 and are therefore not displayed.

4.3   Diagnostics

Most standard diagnostic tests are based on a normal error distribution, which renders these tests unfeasible for our purposes due to the use of the gamma distribution. Also, as our model specification has two mixture components and switching between the regimes is random, there is no straightforward way to obtain residuals.

In order to investigate the goodness-of-fit of our model, diagnostic evaluations can nevertheless be conducted by means of so-called probability integral transforms of the data. This method was suggested by Diebold et al. (1998) and extended to the multivariate case by Diebold et al. (1999). The probability integral transform in the univariate case (for one IV series) is obtained as

$$z_t = \int_0^{y_t} f_{t-1}(u)du \tag{2}$$

where $f_{t-1}(\cdot)$ is the conditional density of the implied volatility with the chosen model specification. The transforms are independently and identically uniformly distributed

in the range [0,1] if the model is correctly specified. Although commonly employed in the evaluation of density forecasts, this method is also applicable to the evaluation of in-sample fit. In the bivariate case, Diebold et al. (1999) recommend evaluating four sets of transforms: z_t^C, z_t^P, $z_t^{C|P}$, and $z_t^{P|C}$. The transforms z_t^C and z_t^P are based on the marginal densities of the call and put implied volatilities, respectively. Similarly, $z_t^{C|P}$ is based on the density of call IV conditional on put IV, and vice versa for $z_t^{P|C}$.

Graphical analyses of the probability integral transforms are commonplace. These involve both a histogram of the transforms, that allows for determining uniformity, as well as autocorrelation functions of demeaned probability integral transforms and their squares. The graphical approach allows for easily identifying where a possible model misspecification arises. Figure 4 presents the 25-bin histogram and autocorrelations for $z_t^{C|P}$, and Figure 5 for $z_t^{P|C}$ for the unrestricted model. Figures 6 and 7 present the equivalent graphs for z_t^C and z_t^P, respectively.

Most columns of the histograms fall within the 95 % confidence interval, which is based on Pearson's goodness-of-fit test. Although there are some departures from the confidence bounds (between zero and four, depending on the case), there is no indication that the model would not be able to capture the tails of the conditional distribution properly. It must be noted that Pearson's test statistics and confidence interval are not exactly valid, as their calculation does not take estimation error into account. However, this omission most likely leads to rejecting too frequently.

The autocorrelations of the demeaned probability integral transforms also provide encouraging evidence, although some rejections do occur at the five percent (but not at the ten percent) level.[6] There clearly seems to be some remaining autocorrelation in the squares of the demeaned probability integral transforms. This same finding has been made previously with univariate models for volatility data (see Ahoniemi (2007) and Lanne (2006, 2007)). A potential explanation is that the model is not quite sufficient in capturing the time-varying volatility of implied volatility.

The removal of dummy variables from the unrestricted model results in a clear deterioration in the autocorrelation diagnostics, and consequently, we have deemed the inclusion of weekly seasonality effects relevant for our model. The diagnostics for the restricted model, or the model without cross effects, are somewhat better than those for the unrestricted model, especially where autocorrelations are concerned.[7] The improvement in diagnostics due to the removal of cross effects is surprising, as the cross terms are statistically significant and improve forecasts (see Section 6 for discussion on forecasts).

In order to verify that our unrestricted model takes the high persistence in the data into account, we compare the autocorrelation functions (ACF) estimated from the call and put IV data to those calculated from data simulated with our model. Figure 8 depicts the autocorrelation functions of NIKC and NIKP, as well as the autocorrelation functions generated by the unrestricted mixture-BVMEM model after simulating 100,000 data points. A 95% confidence band is drawn around the estimated autocorrelation functions. The band is obtained by simulating 10,000 series of 3,194 data points (equal to the full sample size), and forming a band that encompasses 95% of the autocorrelations

---

[6]The confidence bands of the autocorrelations are also calculated without estimation error accounted for.

[7]To save space, the diagnostic graphs for the unrestricted model without dummy variables and for the restricted model are not presented in the paper, but are available from the authors upon request.
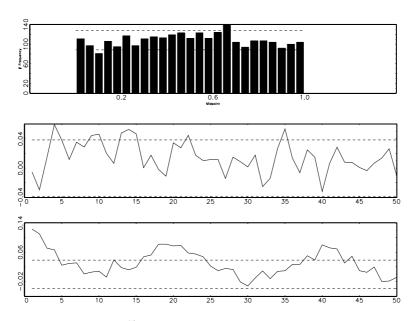
Figure 4: Diagnostic evaluation of $z_t^{C|P}$: NIKC conditional on NIKP. Histograms of probability integral transforms in the upper panel, and autocorrelation functions of demeaned probability integral transforms (middle panel) and their squares (lower panel). The dotted lines depict the boundaries of the 95% confidence interval.
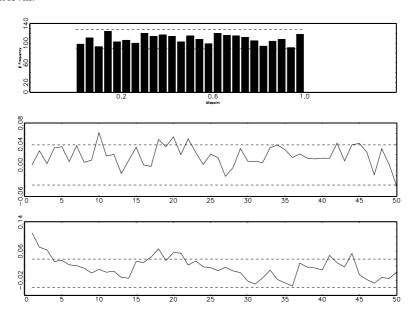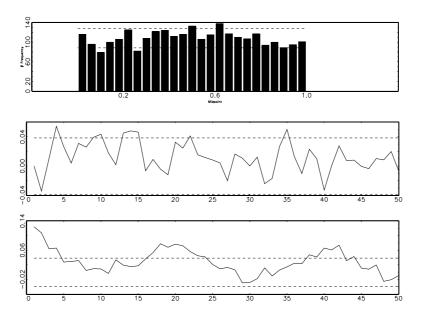

Figure 5: Diagnostic evaluation of $z_t^{P|C}$: NIKP conditional on NIKC. Histograms of probability integral transforms in the upper panel, and autocorrelation functions of demeaned probability integral transforms (middle panel) and their squares (lower panel). The dotted lines depict the boundaries of the 95% confidence interval.

14

Figure 6: Diagnostic evaluation of z_t^C: Marginal NIKC. Histograms of probability integral transforms in the upper panel, and autocorrelation functions of demeaned probability integral transforms (middle panel) and their squares (lower panel). The dotted lines depict the boundaries of the 95% confidence interval.
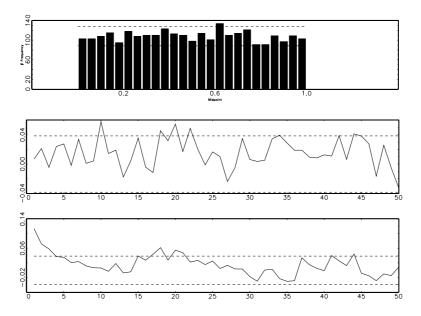

Figure 7: Diagnostic evaluation of z_t^P: Marginal NIKP. Histograms of probability integral transforms in the upper panel, and autocorrelation functions of demeaned probability integral transforms (middle panel) and their squares (lower panel). The dotted lines depict the boundaries of the 95% confidence interval.

pointwise at each lag. As the ACFs generated by our model fall within the band at each lag, it can be concluded that the observed ACFs could have been generated by our mixture-BVMEM model.
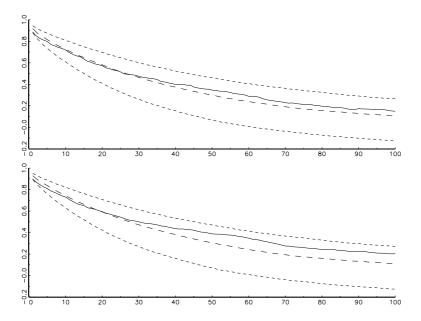

Figure 8: NIKC (upper panel) and NIKP (lower panel) autocorrelation functions. The solid lines depict the ACFs estimated from the full sample of the data, the lines with long dashes are the ACFs implied by the mixture-BVMEM model with 100,000 simulated data points, and the lines with short dashes draw 95% confidence bands around the ACFs.

The diagnostics underscore the necessity for using a mixture model in this case. We also estimated a BVMEM model with only one regime, and the diagnostic checks clearly reveal its inadequacy. In particular, the four histograms for that model show that this specification fails to account for the tails of the conditional distribution, giving too little weight to values close to zero and unity and too much weight to the mid-range of the distribution.[8] However, the imbalance in the histograms is not as severe as with the univariate models in Ahoniemi (2007), thus indicating that even for models without a mixture structure, joint modeling improves the fit to the Nikkei 225 IV data somewhat.

5   Impulse Response Analysis

The bivariate nature of our model allows for a further analysis of how the variables adjust dynamically to shocks. In order to investigate this issue, impulse responses of various types are calculated with both model specifications presented above. This analysis should also uncover more evidence pertaining to the persistence of the data. The more interesting specification for this purpose is naturally the unrestricted model, which includes lagged cross terms in the first regime. Also, as the coefficients of the cross terms are significant, the unrestricted model specification is favored over the restricted

---

[8] The estimation results and diagnostic evaluation for the no-mixture model are available from the authors upon request.

version.[9]

We generate the impulse responses by simulating data according to the conditional mean profiles method proposed by Gallant et al. (1993). It turns out that after approximately 40 periods, the effects of all considered shocks go to zero. Therefore, we present impulse responses up to 40 periods (trading days) ahead. The calculation of the impulse response functions proceeds as follows: we generate 1,000 series of 40 random error terms from gamma distributions with the shape and scale parameters estimated above. Also, we generate 1,000 series of 40 random numbers that are uniformly distributed on the interval [0,1]. These series are used in each period to determine which regime the model is in: if the value of the random number exceeds the value of π, the mean equation for the second regime is used. To get initial values, a starting point in the data set is chosen, and then 1,000 paths, forty days ahead into the future, are simulated from that point onwards with the random error terms, random regime indicators, and estimated parameter values. Another set of 1,000 paths are also simulated, this time with a shock added to the values of NIKC, NIKP, or both in time period 0. The baseline value and the value affected by the shock are calculated simultaneously, so that the same random error terms and regime indicators are used for both. The averages of the 1,000 realizations are taken for each of the forty days, and the impulse response function is then obtained as the difference between the series affected by the shocks and the baseline series without the shocks.

In order to select a realistic magnitude for the shocks, we follow Gallant et al. (1993) and study a scatter plot of demeaned NIKC and NIKP. The scatter plot, shown in Figure 9, helps to identify perturbations to NIKC and NIKP that are consistent with the actual data. As expected, the scatter plot reveals a strong correlation in the two time series. On the basis of the graphical analysis, six different plausible shock combinations are selected: (10,10), (10,0), (0,10), (-10,-10), (-10,0) and (0,-10). In other words, the shock is introduced directly into the value of NIKC or NIKP (or both), rather than into the error terms of the model.

The impulse responses for the first three shock combinations are presented in Figures 10, 11, and 12, respectively. The starting point in the data was July 11, 1996, a time when both NIKC and NIKP were historically quite low. Four noteworthy conclusions can be drawn from the analysis. Most importantly, put-side IV recovers from shocks more rapidly than call-side IV, which is evident in all three figures - even when the shock affects only put-side IV (Figure 12). This result could be based on the phenomenon documented by Bollen and Whaley (2004): the demand for ATM index puts drives the level of ATM implied volatility (in U.S. markets). Also, trading volumes for puts are higher (measured with number of contracts). Therefore, the pricing of puts may be somewhat more efficient, allowing shocks to persist for shorter periods of time than in a less efficient market.

Second, the effects of shocks take a relatively long time to disappear entirely: some thirty trading days, or six weeks, seem to elapse before the effect of a shock is completely wiped out. This finding gives further support to the existence of considerable persistence in the data.

Third, the impulse responses are similar regardless of the starting point that is selected from the data. Four different starting points were in fact considered: a moment

---

[9]Dummy variables are removed from the mean equations in the impulse response analysis. The removal of weekly seasonality does not affect the general shape of the impulse response functions, but makes results more easily readable from graphs.
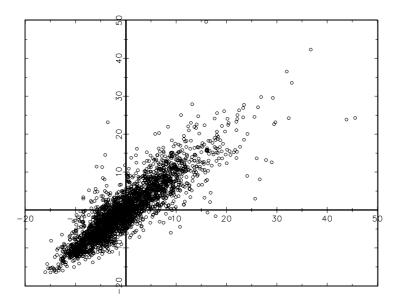
Figure 9: Scatter plot of demeaned NIKC (x-axis) and demeaned NIKP (y-axis) for in-sample period (1.1.1992 - 31.12.2002).

when both IVs were low, a moment when both were high, a moment when NIKC was considerably higher than NIKP, and a moment when NIKP was considerably higher than NIKC. This third result is to be expected as the nonlinearity in the mixture-BVMEM model arises primarily through the mixture of two regimes, with the selection of the mixture component being random rather than dependent on the past values of the implied volatilities. The fourth and final observation is that our model does not allow positive and negative shocks to have effects of differing magnitude. Therefore, only the impulse responses to positive shocks are presented.

Without the evidence on historical values provided by the scatter plot, it could be argued that a shock of the type (10,0), or any shock with a clearly different magnitude for call and put IVs, is not realistic. As outlined above, both IVs represent the market's expectation of future volatility, and should thus be equal. However, empirical analyses again lend support to the fact that call and put IV can differ even considerably at times, due to market imperfections and demand shocks. As an example, the difference between NIKP and NIKC is greatest on Sept. 12, 2001, or immediately after the 9/11 terrorist attacks, when the demand for put options was extremely high. On that day, the difference between the put and call implied volatilities was 33.6.

With the restricted model, or the model without cross effects, the impulse responses look very different. The effect of a shock lasts for less than ten days, or less than two weeks. As there are no lagged cross terms in this model specification, if a shock affects only one variable, the other is (naturally) entirely unaffected and the impulse response is flat.

6 Forecasts

In this section, we turn our attention to the forecasting ability of the two models outlined in Section 4.2. Forecast evaluation is based on two separate criteria: the direction of
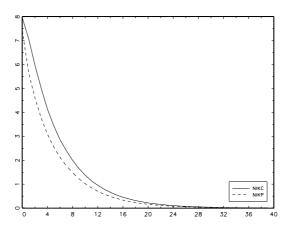
Figure 10: Impulse response function for a shock of (10,10) with the unrestricted model.
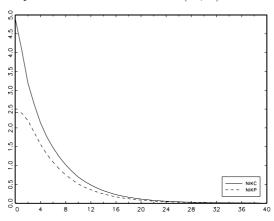

Figure 11: Impulse response function for a shock of (10,0) with the unrestricted model.
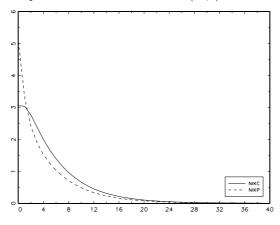

Figure 12: Impulse response function for a shock of (0,10) with the unrestricted model.

change in IV as well as the traditional forecast accuracy measure mean squared error (MSE). It is of particular interest whether the inclusion of cross effects between the two time series can improve the earlier forecast performance of univariate models for NIKC and NIKP investigated by Ahoniemi (2007).

Both daily (one-step-ahead) and five-step-ahead forecasts were calculated with the mixture-BVMEM model specifications outlined above for the 486-day out-of-sample period of 1.1.2003 - 31.12.2004. Days when public holidays fall on weekdays and the observed value of implied volatility does not change were omitted from the data set. Parameter values are treated in two ways: they are either estimated once using the data from the in-sample period and then kept fixed, or re-estimated each day. If parameter values are not stable over time, there can be added value in updating them before calculating each new forecast. When parameter values are updated daily, the forecasts are calculated from rolling samples. In other words, the first observation is dropped and a new one added each day, in order to include information that is as relevant as possible.

The one-step-ahead forecasts are evaluated in terms of both directional accuracy and MSE. Although an accurate forecast of the future level of IV can be valuable to all market participants with risk management concerns, a correct forecast of the direction of change in implied volatility can be useful for option traders. Various option spreads, such as the straddle, can yield profits for the trader if the view on direction of change (up or down) is correct, ceteris paribus.

The forecast results are summarized in Table 3. The directional accuracy of the bivariate model in the two-year out-of-sample is superior to the performance of univariate models. The BVMEM model predicts the direction of change correctly on 348 days out of 486 for NIKC, and on 351 days for NIKP. This is in contrast to the results in Ahoniemi (2007), where the best figures from multiplicative models were 336 and 321 for NIKC and NIKP, respectively. There would appear to be some value to updating parameter values each day. This improves directional accuracy for NIKC clearly, and yields a lower mean squared error for both series of forecasts. However, the direction of change is predicted correctly for NIKP on one day more when daily updating is not employed. Both the unrestricted and the restricted model make more upward mistakes for NIKC, i.e. the models make a prediction of an upward move too often. When predicting the direction of change of NIKP, the unrestricted model forecasts a move downwards too often, but the restricted model a move upwards too often.

A useful statistical test of the sign forecasting ability of the BVMEM models is the test statistic presented in Pesaran & Timmermann (1992). This market timing test can help confirm that the percentage of correct sign forecasts is statistically significant. The p-values from the test are below 0.00001 for all the series of forecasts in Table 3, so the null hypothesis of predictive failure can be rejected at the one-percent level for all four forecast series.

The values for MSE in Table 3 indicate that more accurate forecasts can be obtained for NIKC. Mean squared errors are lower than with the univariate models in Ahoniemi (2007), even with the restricted model. The Diebold-Mariano (1995) test (henceforth the DM test) confirms that the improvement upon the equivalent univariate models is statistically significant for the unrestricted model, with the null hypothesis of equal predictive accuracy rejected at the five-percent level for NIKC and the one-percent level for NIKP. Again, this lends support to the joint modeling of the time series, and the inclusion of cross effects. The forecast accuracy of NIKC with updating coefficients is significantly better than that with fixed coefficients. For NIKP, the difference between

the two alternative treatments of parameter values is not statistically significant.[10]

	NIKC			NIKP		
	Correct sign	*%*	*MSE*	*Correct sign*	*%*	*MSE*
Unrest., updating	**348**	**71.6%**	**4.21**	350	72.0%	**5.24**
Unrest., fixed	341	70.2%	4.31	**351**	**72.2%**	5.26
Rest., updating	332	68.3%	4.34	336	69.1%	5.49
Rest., fixed	332	68.3%	4.39	332	68.3%	5.55

Table 3: Correct sign predictions (out of 486 trading days) and mean squared errors for forecasts from the BVMEM model with both updating and fixed parameter values. The best values within each column are in boldface.

Overall, the results obtained for the Nikkei 225 index option market are superior to those obtained for e.g. the U.S. market. Ahoniemi (2006) finds that ARIMA models can predict the correct direction of change in the VIX index on 62 percent of trading days at best with an identical out-of-sample period. Brooks and Oozeer (2002) model the implied volatility of options on Long Gilt Futures that are traded in London. Their model has a directional accuracy of 52.5 %. Pesaran and Timmermann (1995) predict the sign of excess returns in the U.S. stock market, with results falling within the range of 58% to 60.5%. Gençay (1998) uses a technical trading strategy for the Dow Jones Industrial Average and achieves the correct directional forecast on 57-61 percent of trading days. Our earlier discussion on the effects of limits to arbitrage could perhaps explain why Japanese IV is more predictable in sign than the IV in other markets. If arbitrage is more difficult to carry out in Japan, option prices can depart from their true values to a greater degree, making the market more forecastable.

Table 4 presents the MSEs for the 482 five-step-ahead forecasts that could be calculated within the chosen out-of-sample. The unrestricted model continues to be the better forecaster for NIKP, but surprisingly, the simpler model, or the specification without cross effects, yields lower MSEs for NIKC. The Diebold-Mariano test also rejects the null of equal forecast accuracy at the ten-percent level when comparing the restricted and unrestricted models with updating coefficients for NIKC, but not at the five-percent level. For the corresponding NIKP values (8.79 and 9.36), the null is not rejected. The results for NIKP are better than in Ahoniemi (2007), but for NIKC, the univariate models provide lower mean squared errors. The DM test does not reject the null when the best MSEs from univariate and bivariate models are compared (this applies to both NIKC and NIKP). Therefore, no conclusive evidence is provided regarding the best forecast model for a five-day horizon, but in statistical terms, the BVMEM model is at least as good as univariate models.

---

[10]A bivariate model specification without cross effects and dummy terms is not a better forecaster than univariate models, regardless of whether directional accuracy or MSE is used as the measure of forecast performance. For example, the directional accuracy of this model is 332 out of 486 at best for NIKC, and 318 for NIKP. The detailed results for this third model specification are available from the authors.

	NIKC	NIKP
Unrestricted Model, updating	7.21	8.79
Unrestricted Model, fixed	7.65	**8.78**
Restricted Model, updating	**6.48**	9.36
Restricted Model, fixed	6.60	9.43

Table 4: Mean squared errors for 482 five-step-ahead forecasts from the BVMEM model with both updating and fixed parameter values. The best values within each column are in boldface.

7 Conclusions

It has often been empirically observed that implied volatilities calculated from otherwise identical call and put options are not equal. Market imperfections and demand pressures can make this phenomenon allowable, and this paper seeks to answer the question of whether call and put IVs can be jointly modeled, and whether joint modeling has any value for forecasters.

We show that the implied volatilities of Nikkei 225 index call and put options can be successfully jointly modeled with a mixture bivariate multiplicative error model, using a bivariate gamma error distribution. Diagnostics show that the joint model specification is a good fit to the data, and coefficients are statistically significant. Two mixture components are necessary to fully capture the characteristics of the data set, so that days of large and small shocks are modeled separately. There are clear linkages between the implied volatilities calculated from call and put option prices, as lagged cross terms are statistically significant. The IV derived from put options is a more important driving factor in our model than the IV from calls, as dummy variables for Friday effects of put-side IV are revealed to be significant and to improve the diagnostics of the joint model.

Impulse response analysis indicates that put-side IV recovers more quickly from shocks than call-side IV. Shocks persist for a relatively lengthy period of time (thirty trading days), which is consistent with good forecastability. Also, as the nonlinear feature of our model is primarily the random switching between regimes, the point of time in which a shock is introduced does not affect the behavior of the impulse response functions.

The BVMEM model provides better one-step-ahead forecasts than its univariate counterparts. Both directional accuracy and mean squared errors improve when jointly modeling call and put implied volatility. The direction of change in implied volatility is correctly forecast on over 70% of the trading days in our two-year out-of-sample period. When forecasting five trading days ahead, the BVMEM model is at least as good as univariate models in statistical terms. Based on the combined evidence from all forecast evaluations, we conclude that joint modeling and the inclusion of cross effects improves the forecastability of Nikkei 225 index option implied volatility, and can provide added value to all investors interested in forecasting future Japanese market volatility.

References

[1] Ahoniemi, K. (2006). Modeling and Forecasting Implied Volatility - an Econometric Analysis of the VIX Index. HECER Discussion Paper, 129.

[2] Ahoniemi, K. (2007). Multiplicative Models for Implied Volatility. HECER Discussion Paper, 172.

[3] Bates, D.S. (1991). The Crash of '87: Was It Expected? The Evidence from Options Markets. The Journal of Finance, 46, 1009–1044.

[4] Bollen, N.P.B. & Whaley, R.E. (2004). Does Net Buying Pressure Affect the Shape of Implied Volatility Functions?. The Journal of Finance, 59, 711–753.

[5] Brooks, C., Oozeer, M.C. (2002). Modeling the Implied Volatility of Options on Long Gilt Futures. Journal of Business Finance & Accounting, 29, 111–137.

[6] Buraschi, A. & Jackwerth, J. (2001). The Price of a Smile: Hedging and Spanning in Option Markets. The Review of Financial Studies, 14, 495–527.

[7] Chan, K.C., Cheng, L.T.W. & Lung, P.P. (2004). Net Buying Pressure, Volatility Smile, and Abnormal Profit of Hang Seng Index Options. The Journal of Futures Markets, 24, 1165–1194.

[8] Cipollini, F., Engle, R.F. & Gallo, G.M. (2006). Vector Multiplicative Error Models: Representation and Inference. NBER Working Paper, 12690.

[9] Das, S.R. & Uppal, R. (2004). Systemic Risk and International Portfolio Choice. The Journal of Finance, 59, 2809–2834.

[10] Davies, R.B. (1977). Hypothesis Testing When a Nuisance Parameter is Present Only Under the Alternative. Biometrika, 64, 247–254.

[11] Dennis, P., Mayhew, S. & Stivers, C. (2006). Stock Returns, Implied Volatility Innovations, and the Asymmetric Volatility Phenomenon. Journal of Financial and Quantitative Analysis, 41, 381–406.

[12] Diebold, F.X., Gunther, T.A., & Tay, A.S. (1998). Evaluating Density Forecasts with Applications to Financial Risk Management. International Economic Review, 39, 863-883.

[13] Diebold, F.X., Hahn, J, & Tay, A.S. (1999). Multivariate Density Forecast Evaluation and Calibration in Financial Risk Management: High-frequency Returns on Foreign Exchange. The Review of Economics and Statistics, 81, 661–673.

[14] Diebold, F.X. & Mariano, R.S. (1995). Comparing Predictive Accuracy. Journal of Business & Economic Statistics, 13, 253–263.

[15] Dumas, B., Fleming, F. & Whaley, R.E. (1998). Implied Volatility Functions: Empirical Tests. The Journal of Finance, 53, 2059–2106.

[16] Ederington, L. & Guan, W. (2002). Why Are Those Options Smiling?. Journal of Derivatives, 10:2, 9–34.

[17] Engle, R.F. (2002). New Frontiers for ARCH Models. Journal of Applied Econometrics, 17, 425–446.

[18] Engle, R.F. & Gallo, G.M. (2006). A Multiple Indicators Model for Volatility Using Intra-daily Data. Journal of Econometrics, 131, 3–27.

[19] Figlewski, S. (1989). Options Arbitrage in Imperfect Markets. The Journal of Finance, 44, 1289–1311.

[20] Fleming, J. (1999). The Economic Significance of the Forecast Bias of S&P 100 Index Option Implied Volatility. In: Boyle, P., Pennacchi, G. & Ritchken, P. (Eds), Advances in Futures and Options Research, Vol. 10. Stamford, Conn: JAI Press, 219–251.

[21] Gallant, A.R., Rossi, P.E. & Tauchen, G. (1993). Nonlinear Dynamic Structures. Econometrica, 61, 871–907.

[22] Garleanu, N.B., Pedersen, L.H. & Poteshman, A.M. (2006). Demand-based Option Pricing. CEPR Discussion Paper, 5420.

[23] Gençay, R. (1998). Optimization of Technical Trading Strategies and the Profitability in Securities Markets. Economics Letters, 59, 249–254.

[24] Harvey, C.R. & Whaley, R.E. (1992). Market Volatility Prediction and the Efficiency of the S&P 100 Index Option Market. Journal of Financial Economics, 31, 43–73.

[25] Hentschel, L. (2003). Errors in Implied Volatility Estimation. Journal of Financial and Quantitative Analysis, 38, 779–810.

[26] Jackwerth, J.C. (2000). Recovering Risk Aversion from Option Prices and Realized Returns. The Review of Financial Studies, 13, 433–451.

[27] Lanne, M. (2006). A Mixture Multiplicative Error Model for Realized Volatility. Journal of Financial Econometrics, 4, 594–616.

[28] Lanne, M. (2007). Forecasting Realized Exchange Rate Volatility by Decomposition. International Journal of Forecasting, 23, 307–320.

[29] Lehmann, B.N. & Modest, D.M. (1994). Trading and Liquidity on the Tokyo Stock Exchange: A Bird's Eye View. The Journal of Finance, 49, 951–984.

[30] Liu, J. & Longstaff, F.A. (2004). Losing Money on Arbitrage: Optimal Dynamic Portfolio Choice in Markets with Arbitrage Opportunities. The Review of Financial Studies, 17, 611–641.

[31] Mayhew, S., Stivers, C. (2003). Stock Return Dynamics, Option Volume, and the Information Content of Implied Volatility. The Journal of Futures Markets, 23, 615–646.

[32] Mo, H. & Wu, L. (2007). International Capital Asset Pricing: Evidence from Options. Journal of Empirical Finance, 14, 465–498.

[33] Nagao, M., & Kadoya, M. (1970). The Study on Bivariate Gamma Distribution and its Applicability (1). Annals of the Disaster Prevention Research Institute of Kyoto University 13B, 105–115 (in Japanese with English synopsis).

[34] Nelson, D.B. & Cao, C.Q. (1992). Inequality Constraints in the Univariate GARCH Model. Journal of Business & Economic Statistics, 10, 229–235.

[35] Noh, J., Engle, R.F. & Kane, A. (1994). Forecasting Volatility and Option Prices of the S&P 500 Index. Journal of Derivatives, 2, 17–30.

[36] Peña, I., Rubio, G. & Serna, G. (1999). Why do we smile? On the Determinants of the Implied Volatility Function. Journal of Banking & Finance, 23, 1151–1179.

[37] Pesaran, M.H. & Timmermann, A.G. (1992). A Simple Non-parametric Test of Predictive Performance. Journal of Business & Economic Statistics, 10, 461–465.

[38] Pesaran, M.H., Timmermann, A.G. (1995). Predictability of Stock Returns: Robustness and Economic Significance. The Journal of Finance, 50, 1201–1228.

[39] Poon, S-H. & Pope, P.F. (2000). Trading Volatility Spreads: a Test of Index Option Market Efficiency. European Financial Management, 6, 235–260.

[40] Rubinstein, M. (1994). Implied Binomial Trees. The Journal of Finance, 49, 771–818.

[41] Shleifer, A. & Vishny, R.W. (1997). The Limits of Arbitrage. The Journal of Finance, 52, 35–55.

[42] Yue, S., Ouarda, T.B.M.J. & Bobée, B. (2001). A Review of Bivariate Gamma Distributions for Hydrological Application. Journal of Hydrology, 246, 1–18.